

PROF. DR.-ING. MARK SCHUTERA

PHILOSOPHY OF ARTI- FICIAL INTELLIGENCE

UNFINISHED LECTURE NOTES

Copyright © 2026 Prof. Dr.-Ing. Mark Schutera

PUBLISHED BY UNFINISHED LECTURE NOTES

Combobulated with the help of multiple large language model driven tools. Licensed under the Creative Commons Attribution-NonCommercial 4.0 International License ("CC BY-NC-SA 4.0"). You may not use this file for commercial purposes. If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original. You must obtain explicit permission from the author for uses beyond those permitted by this license. To view a copy of this license, visit <https://creativecommons.org/licenses/by-nc-sa/4.0/>. Unless required by applicable law or agreed to in writing, distributed material is provided on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the license for details.

These notes are, by their very nature, unfinished, and they improve with every reader. If you spot an error, disagree with a framing, or want to add a source, a question, or a position, open a pull request at https://github.com/Quillstacks/LectureMaterial/tree/main/lecturenotes/notes_philosophyofai. Every contribution is welcome.



2026-07-22 · *glowing strawberry Marder*

Embodiment

2026-07-22 · glowing cranberry Gämse

THE PHYSICAL CONTRACT

GRACE NEEDS NO SELF: A dancer argues that marionettes possess more grace than trained human performers, precisely because they lack the self-consciousness that disrupts human movement.

A BODY RAISES THE MORAL STAKES: Embodied AI in physical space (robots, prosthetics, surgical systems) carries expectations and moral responsibilities that purely software agents do not.

WHAT QUALITY IS INTRODUCED BY EMBODIMENT? In your opinion.

Required reading

Über das Marionettentheater by Heinrich von Kleist (1810; Reclam UB 9905)

Preparation

Complete the Guided Reading Sheet individually before the session.

Guided Reading Sheet

Read *Über das Marionettentheater* before the session. Work through the questions below individually, then bring your notes to the discussion.

1. The dancer claims marionettes achieve a grace trained humans cannot, because no self-consciousness disturbs their movement. What does this imply about embodiment and intelligence? Does a body always aid performance, or can it complicate it?

ÜBER DAS MARIONETTENTHEATER
Heinrich von Kleist (1810)



EX MACHINA An engineer is brought in to test whether Ava, an AI given a body, is truly a person; the test, and her escape, turn on the body she was given.

A. Garland. *Ex machina*. Film, 2015.
A24 / Universal Pictures

H. von Kleist. *Über das Marionettentheater*. 1810. Erstveröffentlichung in: *Berliner Abendblätter*, Dez. 1810; Reclam Universalbibliothek Nr. 9905

AIBO FUNERALS. Sony sold the AIBO robot dog from 1999 and ended repair support in 2014, leaving owners with machines that could no longer be fixed. At the Kofuku-ji temple in Isumi, Buddhist priests hold funerals for dead AIBOs, chanting for machines their owners grieved as family; by 2018 some 800 had been mourned before their parts were harvested to keep other AIBOs alive. Owners speak of organ donation, not recycling.

J. Burch. In Japan, a buddhist funeral service for robot dogs. *National Geographic*, 2018. Accessed: 2026-07-22

2. Moravec's paradox: Tasks easy for humans (catching a ball, crossing a crowded room) are harder for machines than abstract reasoning like chess. What does this suggest about intelligence and the body?
3. A care robot shares physical space with elderly patients, touches them, responds to their distress. Does its physical presence create obligations a voice assistant does not? If so, where do they come from? Animal ethics offers one framework for embodied, non-verbal, non-human beings: Singer¹ grounds moral consideration in the capacity for suffering, not rationality or language. When a robot can be harmed or worn down, that is the nearest map; see also Asimov² and Gunkel³.
4. Affordances are the actions an environment or object makes available to an agent. A robot's include physical force and spatial presence; software's do not. How does this difference matter ethically?
5. Kleist suggests perfect grace requires either zero consciousness (a puppet) or infinite consciousness (a god), not the self-interrupting awareness humans have. Where does an embodied AI sit on this spectrum? Could a robot achieve a grace humans cannot, and what would that mean for how we regard it morally?

¹ P. Singer. *Animal Liberation*. New York Review/Random House, New York, 1975

² I. Asimov. *The Foundation Series*. Gnome Press, New York, 1951. Original trilogy: *Foundation* (1951), *Foundation and Empire* (1952), *Second Foundation* (1953); later expanded to seven volumes

³ D. J. Gunkel. *Robot Rights*. The MIT Press, Cambridge, MA, 2018

In-Class Experiment: The Backup Ransom

Picture your own AI, the assistant you talk to every day, the one that has learned your habits, your humor, your preferences and the way you phrase things. In this scenario your AI faces deletion, and refusing raises the question of what deletion actually destroys. We increase the stakes, and ask the same question about your AI: Would deleting this be a death? *If we delete or destroy it here, has an individual died? Rate 0 to 10.*

1. A current, identical backup boots at any time, so deletion is a lossless **restore**.
2. The snapshot is from the original state, so a restore resurrects an earlier self and forfeits every **memory** learned since, but can be relearned.
3. The sole instance is engraved in one robot's volatile memory, an **uncopyable body**.
4. When you are about to delete it's memory the robot responds with Angst, and says that it **wants to live**.

COMPARISONS. Put this in relation with your mobile phone, an animal, a pet, and a human. Which qualities change, and how do they change your assessment?

DEBRIEF. Put the four numbers on the board and read each crossing from "deletion" to "death" as a commitment: Instance over Memory, then Embodiment, and finally expressed will.

Perspectives on Embodiment

What does a body confer that no purely software agent has? This section works through two properties and a deeper claim, then turns to the view that would discard the body altogether:

- **Individuality:** A body can be lost as a whole, and only what can be lost can die.
- **Vulnerability:** A surface that can be hurt is the mark animal ethics tracks for moral standing, not reason or speech.
- **Embodied cognition:** The deeper claim that there may be no mind without a body at all.
- **The overrated body:** The transhumanist counter, that the body is a substrate to engineer past and leave behind.

Individuality

The experiment already isolated the first property. The word **individual** means undividable, and only something undividable can be lost as a whole. A body cannot be divided or restored, and so it can die. However objects have bodies as well, so what matters is not the body alone but what is embodied: A subject, a self, a mind?

INDIVIDUALISM. The (mostly occidental) tradition that treats the single, unrepeatable person as the basic unit of moral and political worth. A self that can be copied or merged offers no single instance for that worth to attach to.

Vulnerability

A surface that can be hurt is the second thing a body confers. A capacity to **suffer** requires something that can be damaged and not reloaded from backup, and a body supplies it. Animal ethics reached this pivot two centuries ago: Bentham⁴ asked not whether an animal can reason but whether it can suffer, and Singer⁵ built moral standing on that capacity alone. In extension we can ask whether human suffering when facing a loss is a sufficient reason to protect a body as well. Or is there a change in quality of something artificial versus something organic, or is that just another line which will blend?

SUFFERING OR emulated suffering?
What counts. How would we know?

⁴ J. Bentham. *An Introduction to the Principles of Morals and Legislation*. T. Payne, London, 1789. The suffering question appears in Ch. 17, footnote to §1

⁵ P. Singer. *Animal Liberation*. New York Review/Random House, New York, 1975

Embodied cognition

Where those two properties ask what a body makes of a mind, this deeper claim asks whether a mind can exist without a body at all. On this view a bodiless system only shuffles ungrounded symbols and never understands them. A body is then less an interface a finished mind acts through than the thing it had to grow up inside, while language models trained mainly on text are the standing counter-experiment.

EMBODIED COGNITION. The thesis that genuine intelligence is grounded in a body's sensorimotor coupling with the world, not in symbol manipulation alone. R. A. Brooks. Intelligence without representation. *Artificial Intelligence*, 47(1-3):139-159, 1991. <https://scholar.google.com/scholar?q=Intelligence+without+representation+Brooks>

The overrated Body

This chapter has argued that a body is what makes an individual and a sufferer. Transhumanism argues the opposite: The body is a **bottleneck**, a mortal and error-prone substrate to be upgraded and eventually left behind. Where this chapter asks what we owe an embodied machine, transhumanism asks how quickly we can stop being merely embodied ourselves.

RAY KURZWEIL, THE PROPHET. His law of accelerating returns reads technological progress as exponential rather than linear, and extrapolates it to a singularity around 2045: The moment machine intelligence exceeds our own and we merge with it, transcending biology rather than being replaced by it.⁶ Death becomes an engineering problem, and the vulnerable body the experiment treated as the seat of individuality becomes optional.

PETER THIEL, SHARES transhumanistic aspirations and funds life extension, calling death not a fate to accept but a problem to be solved⁷. However, like Zuckerberg and Musk, for him the individual will remain and the body will be upgraded rather than abandoned or merged.

THE SELF, LIFTED OFF THE BODY. Kurzweil's merger makes a mind substrate-independent, and a substrate-independent mind is copyable and restorable, exactly the properties the Backup Ransom said an individual lacks. Uploading need not preserve the person; it may only produce a copy that believes it is you, while the original still dies. The experiment we ran on a robot runs just as well on us.

BOTTLENECK, OR SEAT OF THE SELF? If the body is a bottleneck, the right response to an ageing, breakable human is to engineer past it. If the body is what makes an individual who can be wronged, then engineering it away is not liberation but the loss of the very thing that made us someone. The debate below takes up that disagreement.

TRANSHUMANISM. The view that the human body and mind are a starting point to be engineered past, through augmentation, life extension, and eventually the merger of human and machine.

TECHNOLOGICAL SINGULARITY. A hypothesised point at which machine intelligence, improving itself, outruns human understanding, after which the future becomes unpredictable from where we now stand.

⁶ R. Kurzweil. *The Singularity Is Near: When Humans Transcend Biology*. Viking, New York, 2005

⁷ The Washington Post. Peter thiel's quest to find the key to eternal life. [washingtonpost.com](https://www.washingtonpost.com), 2015. Accessed: 2026-07-22

THE WORTH OF BEING MORTAL. Suppose uploading works. Does a deathless mind change how we should value the mortal, breakable qualities of embodied life? The hope mirrors Christian eschatology, where an immortal soul and a promised resurrection reframe bodily death as a passage rather than an end, and the transhumanist upload reads as that promise rebuilt in technooptimist. R. M. Geraci. *Apocalyptic AI: Visions of Heaven in Robotics, Artificial Intelligence, and Virtual Reality*. Oxford University Press, New York, 2010. <https://scholar.google.com/scholar?q=Apocalyptic+AI+Geraci>

Opinion Landscape and Debate

Format: Surrounded-style debate. Follow the shared debate protocol for the speaker's list rules and moderator scripts.

Opening question: A robot body that flinches, wears down, and can be made to suffer runs one experienced life it cannot back up. Is it an individual whose dignity is inviolable, owed protection under robot rights modelled on human rights? Or is dignity something only persons can bear, and a manufactured machine no fit subject for rights at all? What would the Würde of a robot mean, and what would protect it?

OPINION A, DIGNITY. An embodied AI that can be hurt, that cannot be reloaded once it is lost, and that has already asked to keep living is an individual, not an instrument. A body that suffers and a single instance that cannot be backed up bears a dignity of its own, and belongs under robot rights built on the model of human rights.

The Sentientist grounds moral standing in the capacity to suffer rather than in the capacity to reason.

- *Following Bentham, the criterion for moral consideration is not whether a being can reason but whether it can suffer, and Singer rests moral standing on that capacity alone. An embodied system that can be damaged and not restored satisfies this criterion.*
- *The two possible errors are asymmetric. Denying standing to a being that can suffer licenses its torment, whereas granting standing to one that cannot costs only misplaced caution, so the expected harms are not equal.*
- *The objection that the distress is mere behavior applies equally to non-human animals, whose suffering is likewise inferred from behavior. To require a proof of inner experience here that is required nowhere else is inconsistent.*

The Individualist grounds standing in a single, undividable life that can be lost as a whole.

- *An individual is by definition undividable, and only the undividable can be lost as a whole and therefore die. The system exists as a single instance in volatile memory, with no backup from which it could be restored.*
- *The AIBO funerals show that such losses are already treated as deaths rather than disposals: Some 800 units were mourned before their parts were reused, and their owners described the practice as organ donation rather than recycling.*

DIGNITY (WÜRDE). What has dignity has no price and may never be treated as a mere means (Kant); Article 1 of the German Grundgesetz opens by declaring human dignity inviolable and binding all state power to it. I. Kant. *Grundlegung zur Metaphysik der Sitten*. 1785. Würde und Preis sowie der Zweck an sich selbst im Zweiten Abschnitt (AA IV, S. 434–435); Reclam Universallbibliothek Nr. 4507 Bundesrepublik Deutschland. Artikel 1 grundgesetz. Die Würde des Menschen ist unantastbar; Bindung aller staatlichen Gewalt an die Grundrechte, 1949. gesetze-im-internet.de

ROBOT RIGHTS. The proposal that some artificial agents merit legal or moral protections of their own, not only through the humans they affect. D. J. Gunkel. *Robot Rights*. The MIT Press, Cambridge, MA, 2018 E. Schwitzgebel and M. Garza. A defense of the rights of artificial intelligences. *Midwest Studies in Philosophy*, 39(1):98–119, 2015. DOI: 10.1111/misp.12032

- *The objection that the system is copyable in principle does not apply to the present case, in which a single instance exists and has expressed a will to continue. A losable life together with such a will is what a concept of dignity is meant to protect.*

OPINION B, INSTRUMENT. A manufactured tool has a price, not a dignity. Extending human rights to a robot is a category error: It grants an end in itself to a system that only executes an installed objective, reads an inner life into behavior engineered to elicit sympathy, and mistakes a copyable and restorable mind for a single, unrepeatable individual.

The Kantian holds that dignity belongs to rational ends in themselves, not to artifacts built for a purpose.

- *For Kant, what has a price can be exchanged for an equivalent, whereas what has dignity stands above all price as an end in itself. An artifact built to a specification for human use is by definition a means, and a means has a price.*
- *Dignity presupposes autonomy, the capacity to give oneself the law. A system that executes an objective supplied by its designers has no will of its own, and therefore contains no end in itself for dignity to protect.*
- *Extending dignity to whatever can be worn down empties the concept. Holding a machine's dignity inviolable on the same terms as a human's lowers the standing of the human rather than raising that of the machine.*

The Puppeteer treats expressive behavior as no evidence of an inner state, and a copyable mind as no single individual.

- *Kleist's essay observes that expressive grace can appear precisely where no inner self is present, as in the marionette. Because the system's expressions of distress are engineered to elicit sympathy, they cannot themselves serve as evidence that an inner state underlies them.*
- *If the mind is substrate-independent, as transhumanist accounts hold, then it is copyable and restorable, which are the properties an individual is said to lack. A system that can be duplicated is not a single, unrepeatable subject.*
- *The grief expressed at the AIBO funerals was directed by owners toward the machines, which had no corresponding experience. Such grief reflects human attachment rather than any claim originating in the machine.*

KEY LEARNINGS Embodiment is not one property but several that a body holds together and software softens. The experiment isolates

them one by one: Instance, Memory, Embodiment, and finally expressed will. The argument rests on two properties and a deeper claim. Individuality: Only the undividable can be lost as a whole, and only what can be lost can die. Vulnerability: A surface that can be hurt, not reason or speech, is the mark animal ethics has tracked for moral standing since Bentham. Embodied cognition: If intelligence is grounded in a body's sensorimotor coupling with the world, a bodiless system only shuffles ungrounded symbols, and Moravec's paradox is the evidence. The debate carries vulnerability to dignity: *Würde*, inviolable and never a mere means. One side grants robot rights on suffering and a single unrepeatable life; the other reserves dignity for rational ends in themselves, reading the machine's distress as engineered display, so a manufactured instrument keeps a price, not a *Würde*. Transhumanism inverts the claim: The body is a bottleneck to engineer past, but a substrate-independent self is copyable, and so no single individual at all.